

第5章 智能信息处理技术

1. 智能信息处理技术的核心目标是什么？请简要解释其作用。

答：核心目标：利用计算机网络技术实现数据的传递与资源的共享，缩短信息的传播时间，有助于提高信息收集和处理的效率。

作用：通过自动化与智能化技术(如人工智能、数据挖掘)，提升信息处理效率、增加信息处理规模、增强信息价值，并应用于灾害预防、航空航天、教育、智能建造等领域。

2. 海量数据压缩技术的核心原理是什么？请简要说明如何通过压缩技术提高数据存储和传输效率。

答：数据压缩技术的核心原理是通过去除冗余信息以及有效利用数据的内部结构特征，从而以更少的存储空间或更高的传输效率表示原始数据。

通过压缩技术提高数据存储和传输效率的方法：

- ①减少存储空间：压缩后数据占用的存储空间较小，节约存储资源。
- ②提高传输效率：通过数据压缩技术减少传输的数据量，优化带宽利用，减少传输延迟。
- ③降低数据处理难度：通过去除冗余信息，数据压缩技术可以减少对数据的处理复杂度。

3. 集中式数据压缩技术的优点是什么？请解释其在智能建造中的重要性。

答：优点：

- ①集中式压缩技术的一个主要优点是能够大幅度提高数据存储和传输的效率。
- ②集中式压缩技术还能提升数据分析的实时性和精确性。

重要性：

①通过集中式压缩技术，数据可以在收集、传输到中央处理平台的过程中得到有效压缩，减少了对存储空间和带宽的需求，提升了整体系统的效率，这在建筑项目中，尤其是在大规模和长周期的项目中能够显著降低成本和提高响应速度。

②集中式压缩技术还能保证建筑项目在各个阶段都能获得及时、准确的数据支持。

4. 分布式数据压缩技术与集中式压缩技术相比，在哪些方面有优势？请简要说明。

答：优势：

- ①将大规模的实时数据流在生成源头附近进行压缩，减少数据传输时的带宽需求；
- ②避免网络拥堵和延迟，确保系统的实时性和响应能力；
- ③通过并行处理，充分利用多核和多节点的计算能力，提高数据压缩和处理的速度。

5. 在海量数据压缩中，冗余信息的来源有哪些？如何去除这些冗余信息？

答：冗余信息来源：

①时间冗余：在连续的数据序列中，某些数据点与前一个或后一个数据点高度相似(如视频的连续帧)。

②空间冗余：相邻数据之间存在相似性(如图像中相邻像素的颜色相近)。

③统计冗余：数据中某些模式或符号出现的频率较高(如文本中重复的单词或短语)。

去除方法：

①合并相似数据：对时间或空间冗余，算法将相似数据点合并或仅存储差异部分(如差

分编码)。例如,视频压缩中仅存储帧间变化的部分。

②变换与量化(针对有损压缩):变换技术(如离散余弦变换/DCT)将数据转换到更易压缩的域(如频域),分离主要信息与次要细节。量化将连续值离散化,减少数据位数,舍弃不重要的高频信息(如 JPEG 图像压缩)。

③字典编码:建立字典存储重复模式,用短指针替代长数据(如 LZ77 算法)。适用于文本或具有重复结构的图像数据。

6. 异常数据诊断的基本原则和目标分别是什么?

答:基本原则:

①数据分布的对比:异常点通常是与整体数据分布差异较大的点。通过对数据集的统计特性(如均值、标准差等)进行分析可知,异常点往往偏离这些统计特征。在已知数据集的分布形态(如正态分布、均匀分布等)下,异常点通常会表现出较高的偏离度。

②局部与全局的视角:异常点有时是在局部范围内偏离正常数据模式(局部异常),有时则是全局范围内的显著偏离。异常数据诊断不仅要考虑全局的分布情况,还要注意局部区域内的异常。

③数据之间的关系:异常检测不仅仅依赖单一数据点的分布,还需要考虑数据点之间的相互关系。例如,时间序列中的突变、空间数据中的异常点等都可能通过数据之间的依赖性暴露出来。

目标:

①提高数据质量:异常数据可能是由于设备故障、数据录入错误等原因产生的。检测和处理这些异常数据能够增强数据集的质量,从而提高后续数据分析或建模的准确性。

②从异常中发现新的模式或重要信号:异常数据有时并不意味着错误或无意义,相反它有可能揭示了系统中的新变化或重要事件。例如,在工业生产过程中,某个异常点可能预示着设备即将发生故障。通过异常检测,可以及早识别这些重要信号并采取预防措施。

③防止不良决策:如果异常数据未被及时发现,它可能导致模型的错误训练或不准确的决策。特别是在机器学习、统计建模等应用实践中,异常数据可能导致分析结果的偏差。因此,及时发现异常数据并进行处理,有助于预防基于错误数据做出的决策。

④减少噪声的影响:在许多实际工程中,噪声通常被识别为异常数据。噪声会影响模型的性能,特别是在基于数据驱动的方法中。通过异常检测,可以剔除噪声数据,从而提高模型的准确性和稳定性。

7. 异常数据修复有哪几种修复方法?

答:主要修复方法:

①机器学习方法:机器学习算法通过学习数据的内在结构和规律来修复缺失值,尤其适用于复杂的、高维度的数据修复任务。如 K 最近邻、决策树/随机森林、深度学习。

②基于时序的修复方法:在建筑工程项目中,尤其是在施工过程中的监测数据往往是时间序列数据。处理这一类异常数据时,需要考虑时序的连续性和趋势。常见的基于时序的修复方法为:时间序列分解和 Kalman 滤波。

③基于 BIM 的数据修复:基于建筑信息模型(BIM)的数据修复在智能建造过程中发挥着至关重要的作用。BIM 不仅包含几何数据,还涉及工程量、施工进度、质量数据以及设备、人员、材料的动态变化。基于 BIM 的数据修复方法主要包括 BIM 数据校验和数据同步。

8. 根据多源数据融合的输入数据的抽象程度可将多源数据融合技术划分为哪几个层次？

答：根据多源数据融合的输入数据的抽象程度可将多源数据融合技术划分为以下三个层次：

- ①目标级融合(后融合)；
- ②特征级融合；
- ③数据级融合(前融合)。

9. 神经网络模型基本结构按信息输入是否反馈，可以分为哪两种？分别描述它们的特点。

答：一般地，神经网络模型基本结构按信息输入是否反馈可以分为两种：前馈神经网络和反馈神经网络。

①前馈神经网络：

在前馈神经网络中，信息从输入层开始输入，每层的神经元接收前一级输入并输出到下一级，直至输出层为止。整个网络信息输入传输过程中无反馈，即任何层的输出都不会影响同级层。

②反馈神经网络：

在反馈神经网络中，神经元不但可以接收其他神经元的信号，而且可以接收自身的反馈信号。和前馈神经网络相比，反馈神经网络中的神经元具有记忆功能，在不同时刻具有不同的状态。反馈神经网络中的信息传播可以是单向也可以是双向传播。